DOCUMENT RESUME

ED 426 460 EA 029 532

AUTHOR Thomas, Sally; Smees, Rebecca
TITLE Dimensions of Secondary School Effectiveness: Comparing the Findings from Four Academic Studies.
INSTITUTION London Univ. (England). Inst. of Education.
PUB DATE 1998-04-00
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Foreign Countries; Instructional Effectiveness; Outcomes of Education; Performance Based Assessment; *Performance Factors; *School Effectiveness; Secondary Education; Student Development; *Student School Relationship
IDENTIFIERS England; Netherlands; *Value Added; Wales

ABSTRACT

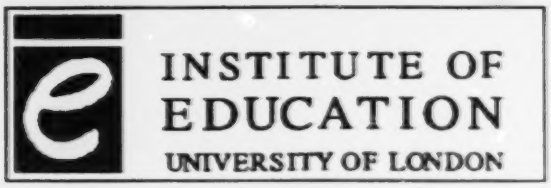
Much research has been devoted to ways in which schools influence students. To study this influence, the identification and definition of dimensions of school effectiveness across a range of outcomes and different geographical and educational policy contexts is presented here. The paper describes the findings from a 3-year study funded by the Economic and Social Science Research Council. The study aimed to extend current knowledge concerning the definition and measurement of secondary school effectiveness by contrasting the findings with new and extended analyses of several independent studies of school and departmental effectiveness. The text establishes the optimal multilevel model for measuring school effectiveness over a set period of time. It compares the optimal models across different geographical areas and educational systems in the United Kingdom (England and Scotland) and also abroad (The Netherlands). The study draws together the findings of these comparative analyses to build definitions of school effectiveness for the UK that encompass a range of different outcomes and also take into account different educational policy contexts. Finally, it addresses how the dimensions of school effectiveness may be operationalized and measured within a school evaluation framework in the UK. (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 460

Draft Working Paper

Not for quotation without prior permission of the authors



Dimensions of Secondary School Effectiveness: comparing the findings from four academic studies

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Sally Thomas
Rebecca Smees

International School Effectiveness & Improvement Centre [ISEIC]
University of London Institute of Education

Paper to be presented at
American Educational Research Association
San Diego
13-17 April 1998

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Thomas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Address for correspondence:

Dr Sally Thomas
ISEIC
Institute of Education
20 Bedford Way
LONDON WC1H 0AL

Tel: ***+44 171 612 6326***
Fax: ***+44 171 612 6344***
Email: ***S.Thomas@ioe.ac.uk***

EA 029 539

BEST COPY AVAILABLE

Summary

This paper describes the findings from a three year Economic and Social Science Research Council (ESRC) funded study[1]. The study aims to extend current knowledge concerning the definition and measurement of secondary school effectiveness by contrasting the findings from new and extended analyses of several independent studies of school and departmental effectiveness. The studies include: the Improving School Effectiveness project funded by the Scottish Office Education and Industry Department (SOEID), the Lancashire local education authority (LEA) value added project, the Department for Education and Employment (DFEE) funded study of 1993-95 General Certificate of Education (GCE) Advanced level results and the Differential School Effectiveness project funded by the ESRC[2]. Using this approach the paper aims to:

(1) To establish the optimal multilevel model (or models) for measuring school effectiveness, over a set period of time, using a value added approach in a range of different pupil outcomes (academic and non-cognitive).

(2) To compare the optimal models across different UK geographical areas (inner city, county LEAs) and education systems within the UK (England, Scotland) and also abroad (Netherlands).

(3) To draw together the findings of these comparative analyses to build a definition of school effectiveness for the UK that encompasses a range of different outcomes and takes into account different educational policy contexts (ie LEAs, regions, education systems).

The major objective of the study is therefore to identify and define the dimensions of school effectiveness (using multilevel techniques) across a range of outcomes and different geographical and educational policy contexts. Issues concerning how the dimensions of school effectiveness may be operationalised and measured within a school evaluation framework in the UK will also be discussed, alongside the current UK government recommendations for a national value added system.

Key Words: school effectiveness, value added, multilevel modelling

[1] ESRC award reference H53627504395

[2] ESRC award reference R000234130

Background

Over the last 30 years, a considerable body of research has been undertaken into the influence of the school. Early work by Coleman, Jenks and others (Coleman *et al* 1966, Jenks *et al* 1972) concluded that family and neighbourhood characteristics have a greater impact on student performance than individual schools. However, subsequent research (eg Rutter *et al* 1979, Mortimore *et al* 1988, Goldstein *et al* 1993) has demonstrated both that schools typically receive variable intakes of students (some schools receive predominantly advantaged and some predominantly disadvantaged groups) and that the outcomes of schooling are not totally determined by their intakes. This has been illustrated by recent research using what has been termed "value added" measures to indicate the effectiveness of individual schools (eg Goldstein, 1997, Gray 1993, Gray *et al* 1996, McPherson 1992, Nuttall *et al* 1989, Sammons *et al* 1994, Thomas *et al* 1993a, 1993b, 1994, Thomas, 1995, Thomas & Mortimore, 1996). Crucially, these value added measures control for the attainment of students on entry to schools (and, if appropriate, other student characteristics such as gender and social class), and are defined in terms of the *relative progress* made by the students within a school (in comparison to students in other schools).

The Value Added Concept

The value added concept rests on the assumption that schools add 'value' to the achievement of their pupils. In educational research the concept of value added has developed over the last decade from the school effectiveness research literature, although it has been used rather differently in other fields such as economics. It is based on the idea of measuring pupil progress, usually in cognitive outcomes such as reading or mathematics attainment during a given period of time. However, the concept can also be applied to non-cognitive outcomes such as pupils' reported attitudes. In order to measure progress baseline and outcome measures are required at the beginning and end of the time period. Of course, as pupils grow older we would expect progress or improvement to be made and average attainment levels to rise. Therefore, researchers use the term value added to refer to the extra value that is added by schools to pupil attainment (or attitudes) over and above the progress or improvement that might be expected in a normative sense. Value added measures thus seek to establish whether pupils in some schools make relatively greater or less progress than those in other schools over a specified period of time. The most effective of schools would be those in which pupil progress exceeds expectations. This is measured by the **residual** value added by the school.

The value of schools' educational quality is, however, broader than what can be measured by attainment in a few specific areas of pupil activity. Therefore, in the context of this paper which focuses largely on pupil outcomes, it is important to note that a comprehensive value added framework for school self evaluation might also encompass measures related to numerous other aspects of a school's mission, processes and outcomes. This broader approach to school self evaluation encapsulates a practical application of Scheeren's (1990) theoretical model of school functioning. In other words, similar data describing inputs, process and outputs is collected about individual schools but the primary purpose is that the information is used directly by school staff to evaluate their educational policy, practice and improvement processes.

The development of Value Added measures

The development of value added measures of school effectiveness has arisen from a variety of sources, rooted in both academic research and policy related issues. First, many school effectiveness studies, in particular those carried out prior to the mid 1980's, were hampered by

the limited statistical techniques available (for a review see Scheerens, 1992, 1997) and did not have access to the recently developed, sophisticated and now widely preferred method of analysis, multilevel modelling (Goldstein, 1987, 1995). Secondly, the requirements of the 1980 Education Act and the 1991 (schools) Bill (section 16) for schools to publish their "raw" public examination results placed a much greater emphasis on the search for fairer and more accurate measures of school performance and this has led to the increasingly widespread and systematic collection of student data by local education authorities (including information about student examination and assessment outcomes and other student and school characteristics, see Hill, 1997 for a review). Consequently, the advances in statistical techniques and data collection has facilitated the development of more accurate and appropriate measures of school effectiveness in the UK.

Individual schools have also addressed the issues of school performance and effectiveness as an aspect of internal evaluations and external inspections (such as those carried out by the local education authority and, at the national level, by the Office for Standards in Education). In recent years, schools and LEAs have employed a wide variety of different procedures using either pupil background factors (such as socio-economic status) or pupil prior attainment data, or both, as well as different levels of sophistication in the analysis (eg employing individual pupil level data or cruder aggregated school level data) (see for example, Gray 1993, Thomas *et al* 1993b, 1994, Thomas & Mortimore, 1996, Hill, 1997). At the national level interim procedures for assessing school effectiveness have been proposed that employ contextual information about pupils and schools but not pupil prior attainment data (Sammons *et al* 1994) but so far no optimal value added model for measuring school effectiveness has emerged. Nevertheless, the Conservative government's white paper (1996) stated explicitly the need for school staff to monitor the quality of the education they provide and any improvements. Moreover, the government funded National Value Added Project has recently made recommendations about the kind of value added information that should be provided to schools via a national system (SCAA, 1994; Fitz-Gibbon, 1995; SCAA, 1995, SCAA 1997, see Panel 1).

Panel 1

EDUCATION WHITE PAPER (DFE, 1996, Page 53)

The Government's priority is to foster the internal will and capacity of schools to generate their own improvement ... staff and governors of every school should feel that it is directly for them to monitor the quality of the education they provide and improve schools ...

THE NATIONAL VALUE ADDED PROJECT:
Report to the Secretary of State (SCAA, 1997, Page 7)

It is recommended that:

- *results to schools are reported in both tabular and graphical forms; showing both the average progress made and the range of results achieved by pupils from any given starting point;*
- *for the purposes of initial feedback to schools, a simple statistical model should be used;*
- *further investigations should be carried out during the Pilot Year and beyond of the various statistical models available, in order to inform future feedback and other uses of the information.*

A National Value Added System

In spite of the government's recent recommendations, the major difficulty of introducing a national system for value added measures is the lack of reliable standardised assessments to measure the prior attainment of pupils entering school. There are no national assessments of pupils entering primary school and at the secondary level the only appropriate national assessments are taken at the end of primary schooling (ie National Curriculum (NC) Key Stage 2), *before* entering secondary school, or half way through (ie at NC Key Stage 3). Moreover, the national curriculum assessments at Key Stages 1-3, which are reported in terms of ten levels, may not differentiate sufficiently between pupils nor be sufficiently reliable for the purpose of measuring value added. Finely differentiated and reliable attainment measures are necessary to describe accurately pupils' starting point. However, if there were to be any development of the national curriculum assessments, the benefits of teacher and standard task assessments in enhancing the quality of teaching and learning would need to be maintained and at the same time complement this with assessments that can be used for the purpose of measuring value added. Current new developments include the requirement of Local Education Authorities (LEAs) to implement a recognised system of baseline testing for five year Odds (SCAA, 1996). Some LEAs, such as Surrey and Hampshire, have already employed baseline assessment to evaluate school effectiveness for the infant or junior phase (Sammons & Smees, 1997a & b).

Utility of New Research Findings

The aim of this paper is to provide new evidence to assist school staff, policy makers and academics in understanding the multi-faceted nature of school effectiveness and the need to evaluate school performance in a more realistic context. The findings will be of *practical* value, to school staff and inspectors by identifying and defining valid and appropriate measures of school effectiveness that can be employed in the processes of school evaluation and the internal monitoring of performance. For example, evidence concerning the nature of the relationship between schools' effects on pupils' academic and affective/social outcomes will inform schools about how outcomes in different areas may interact.

If schools' effectiveness measures are broadly similar in different areas then an overall measure may be useful and valid (in addition to more detailed measures for individual outcomes, subject departments or for different groups of pupils) and would point to a single *underlying* dimension of school effectiveness. Alternatively, the evidence may suggest that two or more different *underlying* dimensions of effectiveness are required to describe the full complexity of school effectiveness. In addition, the comparative multilevel analyses across regions and education systems will assist educational policy makers in understanding how different regional, socio-economic and educational policy contexts (both in the UK and abroad) may influence the size, extent and stability over time of school effectiveness, as well as the impact of particular pupil characteristics (such as gender and social class) and school factors (such as single sex/mixed schooling) across a range of outcomes.

Below, we briefly review the previous evidence available on value added measures of school effectiveness and associated approaches to school self evaluation before describing the findings of the current study.

Comparing Different Value Added Measures: The Optimal Multilevel Model

Recent research, carried out at the secondary level by Thomas & Mortimore (1996), has been able to take the fullest account of students' attainment on entry to school, employing three different

measures of prior attainment as well as a wide variety of student and school background factors (outside the control of the school) in the multilevel analysis of student outcomes at the end of statutory schooling (at age 16 years). This work started in 1992 and the aim has been to develop, as far as possible, the most accurate, appropriate and fair measures of school effectiveness and to feed these results back to schools and the LEA, in confidence, for the purpose of school self evaluation. Several different academic outcomes were analysed (the General Certificate of Secondary Education (GCSE) total score, and scores in English, mathematics and science) and statistical controls were provided for a range of different student intake factors in order to establish a basic model. Four additional models were also presented for comparative purposes. It was found that schools' raw (unadjusted) and value added residuals were similar for some schools but rather different for others, demonstrating the difference between school effectiveness measures and the absolute level of pupil GCSE examination performance. It was also found that the optimal model for the total GCSE score controlled for students' prior attainments in verbal, quantitative, and non verbal cognitive ability tests, their gender, ethnicity, mobility and entitlement to free school meals, and two 1991 Census factors relating to the students home area (% higher education qualifications and % unskilled [RG Group V]). However the school residuals for this model were very strongly correlated (0.92) with the model which employed only prior attainment data. Therefore the findings indicate that while employing prior attainment data provides the best and most appropriate measure of value added, information about other pupil background factors (such as socio-economic status) can fine tune this measure. This evidence is in line with previous research by Willms (1986, 1992) which employed a more limited dataset. However, several researchers have noted that further research is required to investigate a variety of models of school effectiveness and the consistency of the findings across a range of different student outcomes (eg academic, vocational, affective/social), taking into account the stability of the results over time (eg Scheerens, 1992, Scheerens & Bosker, 1997).

A Range of Different Outcomes

The range of student outcomes employed to investigate school effectiveness has been relatively narrow (most research has focused mainly on academic outcomes, Scheerens 1992, Sammons *et al* 1996) and there is an urgent need to widen the scope of school effectiveness research to include additional outcomes in the vocational and affective/social areas. Of the studies which have examined secondary schools' effects on different outcomes most have focused on the performance in the areas of English and mathematics (eg Willms & Raudenbush, 1989, Smith & Tomlinson, 1989, Goldstein *et al*, 1993, Thomas *et al* 1994). The findings have indicated that schools doing well with students in one aspect are not necessarily effective in all aspects. Similar conclusions have been drawn from a recent ESRC study[3] looking at a wider range of outcomes (six GCSE subjects and one overall GCSE measure) concerned with the consistency and stability over time of inner city secondary school effects (Nuttall *et al* 1992, Thomas *et al* 1997a & b, Sammons *et al* 1997). The evidence between departmental and overall results (taking into account three years of data, 1990-92) ranges from *fairly strong* in some cases to *fairly weak* in others (the correlations of schools' effects on GCSE results are all positive ranging from 0.27 to 0.85). These findings are reflected in research carried out abroad in the Netherlands (eg Luyten 1994) and at the post 16 level (Fitz-Gibbon, 1991).

With regard to the related issue of the stability of school effects over time, the importance of this aspect of school effectiveness has been established by several researchers (eg Raudenbush, 1989, Gray *et al* 1993, 1995, 1996; Thomas *et al* 1997a & b). In general, the evidence indicates that

[3] ESRC award reference R000234130.

for most schools performance is broadly similar over time but in some cases schools results can vary substantially indicating either improvement or decline in performance. In this context, it is important to emphasise that 'real' improvement (or decline) in performance, resulting perhaps from a shift in educational policy or practice, can only be identified by examining long-term changes in results over time (Gray *et al*, 1995, 1996). Recently, researchers have noted the importance of examining in detail the performance trends of individual schools and the educational processes associated with different patterns of improvement (Gray, Hopkins and Reynolds, 1998).

Research in the late 1980's also examined the issue of differential school effects for different groups of students (such as high and low attainers, boys and girls or different ethnic groups) and found that an important aspect of a school's effectiveness was whether it was equally effective for *all* pupil groups of pupils (see for example Nuttall *et al* 1989, Smith & Tomlinson, 1989). However, these studies were somewhat limited in the number of schools investigated or the availability of detailed information about the background and prior achievements of the pupil sample. More recent research, employing detailed pupil level data, has confirmed and extended previous findings on this topic by examining differential departmental effects and established that using an overall measure of school or departmental performance may mask important differences in the relative progress made by different pupil groups, particularly those categorised by prior attainment and ethnicity (Thomas *et al* 1997b).

Thus, overall the evidence suggests strongly the need for further evidence about school performance over time and in detail for different pupil groups, not just in terms of total performance but also at department (or subject) level, as well as in other outcome areas (such as vocational and affective/social) in order to describe adequately the variety of school effects.

Regional Comparisons

Few studies have addressed the issue of regional differences in the size, extent and consistency of school effects or the differential impact of pupil and school background characteristics in different regional, socio-economic and educational policy contexts. Evidence of this kind is vital to inform educational policy makers about the influence of local area, regional and national policy and practise. Gray *et al* (1990) has compared the value added estimates for schools in six different LEAs in the UK and found substantial differences between the estimates of school variation (after controlling for student intake) for different regions. However, the conclusions that can be drawn from these comparisons are limited due to differences in the controls employed for student intake (4 LEAs were lacking prior attainment data) and the small size of school samples (30 or fewer schools in 5 LEAs). At the international level Creemers *et al* (1994) has described a comparative study involving 5 countries, focusing mainly on primary mathematics which is part of the on-going International School Effectiveness Research Programme (ISERP). Although this study is severely limited due to the very small samples of schools in each country (12 or fewer) the findings show important differences between countries in the size and extent of school effects after controlling for student intake. Creemers *et al* (1994) underlines the need for further research to investigate systematically the existence and reasons underlying regional and national differences in school effects with larger samples of schools.

Need for Further Research

At present, the evidence of secondary school effects across different outcomes and regional contexts (using multilevel techniques) is sparse. The previous research described above has focused mainly on a limited range of outcomes (ie academic) and few studies have looked at

comparisons between regions (which vary in terms of both educational policy, socio-economic and other regional factors). There is a need to develop this area of research in the UK and to clarify the findings of school effectiveness studies in the wider regional and national context. In other words to investigate the optimal multilevel model (or models) specification for measuring school effectiveness and to study both consistency across outcomes (eg academic, vocational and affective/social), effectiveness for different pupil groups and stability, or instability, over time in different educational contexts. Thus the major objective for further research is to identify and define the dimensions of school effectiveness to reflect the full complexity of school performance. This will involve also considering the role and influence of primary school effectiveness on secondary school performance (see Sammons *et al* 1995, Goldstein & Sammons 1997) and the comparative size and extent of school level variation in comparison to class level variation (see Rowe *et al* 1994, Hill & Goldstein *forthcoming*).

Aims and objectives of a new comparative study

This study aims to extend current knowledge concerning the definition and measurement of secondary school effectiveness in three major ways. The first objective is to establish the optimal model (or models) for measuring school effectiveness using a value added approach in a range of different pupil outcomes (eg academic and affective/social[4]). The optimal model specification is defined as the most efficient (in terms of statistical criteria) and the most appropriate (in terms of employing factors most relevant to school effectiveness). The methodology will employ multilevel modelling (Goldstein 1987, 1995) and will investigate and contrast the impact on student outcomes of different factors outside the control of the school (such as pupils attainment on entry to school, gender, social class) as well as the *interactions* between these factors. The aim will be to compare the *consistency* of the parameter estimates of these models (ie type and significance of predictor variables, size and extent of school effects) for a range of different outcomes. The significance and importance of employing data over several years will be investigated as a key element of the optimal multilevel models, and, where data is available, the influence of primary school effectiveness on secondary school performance and the necessity to control for variation in student outcomes at the level of the classroom will also be examined.

Panel 2: Key Aims

- **To provide a definition(s) of the dimension(s) of school effectiveness that can be measured, operationalised and built into a comprehensive value added framework for school self-evaluation in the UK.**
- **To provide practical as well as theoretical evidence that will assist school staff, policy makers and academics to evaluate school performance in a more realistic context.**

The second objective is to compare the optimal multilevel models and school residuals according to different regional contexts (eg inner city, county LEAs) and education systems within the UK (eg England, Scotland). The comparison of value added models will also include education systems outside the UK (such as in the Netherlands). The third objective is to identify and define the dimensions of school effectiveness across a range of outcomes and different regional, socio-economic and educational policy contexts. Thus, the key aims of the study are to provide a

[4] Originally it was intended to also examine vocational outcomes in this study. However, unfortunately, due to the difficulty of obtaining a representative sample of student's vocational outcomes (such as GVNQs), this aspect of the project will only be discussed in terms of possible future developments of the value added models.

definition(s) of the dimension(s) of school effectiveness that can be built into a comprehensive value added framework for school self evaluation and also to provide convincing evidence that will assist school staff, policy makers and academics to evaluate school performance in a more realistic context (see Panel 2). The research will focus on school effectiveness at the secondary phase of education. However, appropriate comparisons will also be made with the Post 16 and primary phases of education in order to place the effectiveness of secondary schools within the whole context of state educational provision. For the purpose of clarity, first two objectives of this project have been divided into nine detailed research questions and these are shown in Panel 3.

Panel 3: Research Questions

1 **Which explanatory variables are most important to control for in the analysis?**
Prior attainment in curriculum areas (eg language/ mathematics); IQ; background (eg gender); context (eg % low attainment); interactions.

2 **Subject results**
Should a separate analysis be carried out for each curriculum subject as well as an overall measure of student academic outcomes?

3 **Differential effectiveness**
Should the results for different groups of students such as high and low attainers be examined separately?

4 **Variability over time**
Should a separate analysis be carried out for each individual year/cohort or should a more representative sample such as 2 or more consecutive years/cohorts be employed?

5 **Variability over regions**
Should separate analyses be carried out for different regions or LEAs?

6 **Variability over year groups**
Which period(s) of a pupils school career should be employed to examine effectiveness?

7 **How important is it to incorporate the classroom or teacher level in the analysis?**

8 **How important is it to incorporate the continuity of previous school effects?**

9 **Affective/attitudinal outcomes**
Should additional analyses be carried out for affective/attitudinal outcomes?

Samples and Data

Extensive pupil and school level data has been collected over five years (1993-97) from 99 Lancashire secondary schools including students' GCSE results, three measures of attainment[5] on entry to secondary school (at age 11 years) and other student background variables (see Thomas & Mortimore, 1996 for specific details of variables). Additional outcomes comprising students reported attitudes collected in 1996 and 1997 complement the GCSE and student intake data. In addition to the Lancashire dataset, further equivalent datasets are available from different regions within the UK (London 1990-92 and Scotland 1997), and outside the UK (Jersey 1993-95 and the Netherlands 1995). An additional database is also available which includes the complete

[5] National Foundation of Educational Research (NFER) Cognitive Abilities Test (CAT) with sub-tests for verbal, quantitative and non-verbal abilities.

1994 and 1995 GCE A-level outcomes for all English students matched to previous GCSE attainment (1992/3 or earlier). Thus, six complete datasets relating to a variety of regions in the UK and abroad and including student outcomes in different academic areas (eg total performance score and individual subject results in English, science and mathematics), prior attainment data, and student background characteristics were available for the comparative study (see Panel 4). A summary of the outcome and background data recorded for each dataset is shown in Panel 5.

Panel 4: Samples from Six Datasets

(1) **LANCASHIRE [1993-97] 99 Schools; 61,103 students**
Outcomes: GCSE scores and attitude scales

(2) **LONDON [1990-92] 94 schools; 17,850 students**
Outcomes: GCSE scores

(3) **JERSEY [1993-95] 9 schools; 1,849 students**
Outcomes: GCSE scores

(4) **SCOTLAND [1997] 36 schools; 4,500 students**
Outcomes: standard grade examinations and attitude scales

(5) **NETHERLANDS [1995] 256 schools; 8,543 students**
Outcomes: Dutch language, mathematics attainment of 14/15 year Odds

(6) **ENGLAND [1994-95] 2,700 Institutions; about 500,000 students**
Outcomes: A/AS level examinations

Note: each dataset includes, as a minimum requirement, individual pupil records of cognitive outcomes (language and mathematics), prior language attainment and gender. Only the Lancashire and Scottish datasets include pupil attitude data.

Panel 5: Data Employed for Multilevel Analyses

[1] Cognitive and affective outcomes
- ◊ **total cognitive measure**
- ◊ **language (English/Dutch)**
- ◊ **mathematics**
- ◊ **science**
- ◊ **attitude scales (Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support)**

[2] Baseline measures of prior attainment and attitude
- ◊ **language (English/Dutch)**
- ◊ **mathematics**
- ◊ **general ability/IQ**
- ◊ **attitude scales (Engagement, Pupil Culture, Self Efficacy, Behaviour, Teacher Support)**

[3] Student background characteristics
- ◊ **gender**
- ◊ **age**
- ◊ **entitlement to free school meals**
- ◊ **ethnicity**

[4] Context
- ◊ **% low attaining students (approximately bottom 25%)**

Methodology

Four aspects of methodology have been incorporated into the research design. First, the statistical technique of multilevel modelling will be employed (see Paterson & Goldstein 1991 for an introduction). This technique, a generalised form of multiple regression, incorporates the hierarchical nature of the data and allows the variation in pupil outcomes to be examined at different levels within the hierarchy (eg students are clustered within classes, departments, cohorts, schools and LEAs). It employs student level data (thus maintaining the original relationship between the student outcome measures and different student intake variables) and attaches a measure of statistical uncertainty to the individual school residuals so that any apparent differences or similarities in the results can be realistically interpreted. The MLN software developed as a result of the ESRC sponsored Multilevel Models Project is used for the data analyses (Rasbash & Woodhouse, 1995).

Secondly, the methodology of analysing each student outcome involves using prior (or baseline) attainment data (eg the attainment data collected at age 11 for the purpose of UK secondary transfer) which is the most crucial variable to control in measuring the school effect. In addition, the relative importance of different types of prior measures, such as those relating to the curriculum or to underlying abilities, are investigated (this issue has been previously addressed by work carried out by Madaus *et al* 1979, but multilevel modelling was not employed).

Thirdly the samples of schools employed for each analysis is maximised, where possible. The larger the base of comparison, the more robust is the statistical result of multilevel modelling and the more generalisable are the overall findings. It is very important that an appropriate and representative base of comparison is employed to identify accurately the range and extent of school effects in the UK.

Lastly, in order to facilitate the comparison of optimal models across outcomes and regions all student outcome data has been transformed to standardised scores (ie normal scores with a mean of 0 and a standard deviation of 1). With regard to the analysis of student attitude data (see research question 9), pupil attitude scales have first been created by weighting similar questionnaire items using an appropriate statistical technique (LISREL)[6]. In addition all student and school background data has been recoded (where appropriate) into categories that are standard across all datasets.

Summary of Multilevel Analyses

A variety of two (and three) level models have been contrasted in order to identify the optimal multilevel model for examining school and department effectiveness for each dataset (see Panel 6). This approach allows the average fixed effects as well as the random effects associated with the level of the school and the individual student to be examined. In the case of datasets which include more than one cohort of students it is also possible to examine the effects associated with year to year fluctuations in the results. In other words, a three level model is applied which measures separately the random variation in pupil outcomes into that attributable to the school, cohort (or year) and the pupil level. Essentially, the school residuals obtained from a three level analysis over time represent the mean school effects over several cohorts of students. To identify the optimal model for each outcome and region four aspects of the multilevel results have been examined:

---

[6] See Thomas *et al* 1997, 1998 for a description of how the pupil attitude scales were created.

[I] The weighting and significance of different explanatory variables included in the multilevel model (see Panel 6)[7].

[ii] The percentage reduction in the total and school level variation in pupil outcomes by introducing different explanatory variables in the multilevel model (see Panel 6).

[iii] The percentage of total variation attributable to the pupil, cohort and school levels.

[iv] Standard deviation and range of school residuals.

Panel 6: For each region and cognitive outcome measure seven different models incorporating contrasting explanatory variables have been employed:

Model 1	Intercepts only
Model 2	Language prior attainment measure only
Model 3	IQ prior attainment only
Model 4	All prior attainment measures
Model 5	Prior attainment measures and one school context variable (ie % below average attainers in each school)
Model 6	Prior attainment and pupil background characteristics
Model 7	All measures: prior attainment, background, context

An equivalent set of models was employed for the pupil attitude outcomes but are not reported here.

Having established the most appropriate explanatory variables for the optimal model for each outcome measure the following correlational analyses were carried out to examine in more detail the differences in school residuals across a range of pupil outcomes. Where data was available the analyses were repeated for each dataset and the results subsequently compared.

[i] Correlations between school residuals for different academic outcomes (total score, language, mathematics, science) to identify differences in schools' overall and departmental performance.

[ii] Correlations between school residuals for different groups of pupils (categorised by prior attainment, gender, FSM and ethnicity) to identify differences in school and departmental effects for particular pupil groups*.

[iii] Correlations between school residuals for different cohorts of pupils' aged 16 years (categorised by year of taking outcome assessment such as GCSE) to identify differences in school and departmental effects over time.

[iv] Correlations between school residuals for different year groups (categorised, for example, by outcome assessment at Year 11 [age 16] and Year 9 [age 14]) to identify differences in school and departmental effects for different year groups within a school*.

---

[7] Interactions between explanatory variables will also be examined in the final report of the project.

[v] Correlations between school residuals for cognitive and affective outcomes (cognitive outcomes: total score, language, mathematics, science. affective outcomes: engagement with school, pupil culture, self efficacy, behaviour, teacher support) to identify differences in school performance in two areas: cognitive and affective.

This work is still in progress but a summary of the results - to date - is provided in the next section (analyses marked with a * are not yet completed for all datasets). Extended multilevel analyses are also planned to examine the variation in pupil outcomes [i] across classrooms (using the Scottish dataset); [ii] across regions (using the England GCE A-level dataset); and [iii] cross-classified at the primary and secondary levels (using the 1997 lancashire dataset).

Summary of Findings

◊ In line with previous research, the optimal multilevel model for measuring school and departmental effectiveness controlled for pupils' previous attainment (or attitudes), pupil background factors (such as gender) and school context (such as % Band 3).

◊ Confirming and extending the results of previous research, the findings from two or more datasets suggest that at least three dimensions of secondary school effectiveness can be identified in terms of cognitive outcomes for:

(1) separate academic subjects (such as language, mathematics and science);
(2) different groups of students (such as high and low attainers);
(3) different pupil cohorts (such as consecutive GCSE cohorts over time).

◊ The evidence of point (3) above points to the usefulness of value added measures for individual cohorts to examine in detail the improvement (or decline) in value added scores over time, and also to evaluate the performance of a single cohort of students, or the specific factors relating to their outcomes, or both.

◊ However, value added results that reflect the average results of two or more consecutive cohorts may also provide useful summary measures. Using this approach to examining trends in school performance over time will provide results that do not fluctuate dramatically from year to year and could be described as a kind of rolling average of school performance.

◊ There is some variability across regional datasets in the percentage variance explained by different prior attainment and background measures, the percentage variance due to the school and year level and the range and extent of school effects. These findings tentatively suggest that regional context or policy does have an influence on the range and extent of school performance and points to the value of separate regional measures of school performance.

◊ Finally, new evidence has been found concerning the relationship between secondary schools' effects in two areas (academic and affective). The results show that the correlations between schools' residuals in the two areas examined is weak[8]. This finding is important because it suggests that separate dimensions of effectiveness exist that reflect difference aspects of how schools and teachers can influence pupils' attitudes and achievements.

[8] For example, using the Scottish dataset no correlations - except one - were greater than +/- 0.3.

The implications of the Value Added Results for School Self-Evaluation

The results of this study emphasise the need for school staff to analyse data in a more sensitive and detailed way, at a range of levels: individual pupils; various pupil groups; sub-groups; subject level; whole school; regionally and nationally. Panel 7 summarises different issues and approaches to be considered when interpreting and using value added results for the purpose of school self evaluation. Importantly, schools need to collaborate with other schools at the local, regional and national level in order to provide comparative data for value added analyses. Examples of this approach are provided by the work of some LEAs, such as the Lancashire, Hampshire and Surrey LEA value added projects.

Panel 7: Approaches for Interpreting and Using Value Added Results for School Self Evaluation

- Value added results offer a fairer and more meaningful way of presenting school examination results than the raw unadjusted results. However it is important to consider the importance of confidence limits when making any comparisons between schools — if the confidence intervals of two particular schools overlap then there is no *statistically significant* difference between their performance.
- Bear in mind limitations of the methodology for individual schools. How relevant are issues of: measurement error, missing data, data accuracy and the retrospective nature of the data?
- Track changes in results over time to examine real improvements, or random fluctuations in performance, or both, in relation to school improvement initiatives.
- Examine departmental and/or teacher effectiveness versus summary measures of school effectiveness (eg total GCSE performance for the average pupil) and their implications for whole school policies.
- Examine differential effectiveness for different groups of pupils (eg, boys/girls, high/low attainers) and implications for equal opportunities.
- Examine local or regional differences in value added results between schools and the implications for local, regional or national education policy.
- Employ a wider range of value added measures to reflect more fully the aims of schooling (eg, using pupil attitudes and vocational as well as academic outcomes).
- Contrast the results against other types of data available in schools such as information about the views of key groups obtained using for example teacher and parent questionnaires; and
- For individual pupils and specific groups of pupils (such as boys or girls or certain ethnic groups) value added results can provide additional guidance in monitoring and target-setting. However, the results should be used extremely cautiously, particularly for an individual pupil, bearing in mind other information about an individual's particular circumstances, and the fact that past performance does not necessarily predict future performance.

Moving from Measurement of School Effectiveness to School Improvement

The task of linking school effectiveness measures to school improvement starts with the premise that analysis is the start not the end of the process. Monitoring does not by itself improve performance, nor does it provide definite distinctions or comparisons. Therefore it is important that information about school, departmental and classroom effectiveness is continuously contrasted with current policy and practice. For example some secondary schools in Lancashire have used separate value added subject scores for the most and least able pupils to reflect on and evaluate their systems for setting GCSE pupils. Also schools in Northern Ireland involved in the Department for Education Northern Ireland (DENI) funded Raising School Standards Initiative

(RSSI) aim to use value added measures as well as other evidence to evaluate the impact of particular improvement strategies at the school and classroom levels (Thomas & Elliot, 1997a & b).

Future developments in value added research are likely to build on current findings that investigate the relationship between measures of school performance and the *conditions* that appear to enhance, or hinder, school effectiveness in different types of school context. For example, under what circumstances or conditions does the impact of context (such as %FSM) on school performance vary? This approach also requires the use of both qualitative and quantitative data. Research by Sammons and colleagues (Sammons *et al*, 1997) employed both value added methodology to evaluate school performance and also combined this with interview and questionnaire data to investigate factors and processes related to greater departmental and school effectiveness.

Finally and most importantly, further attention is required on the crucial issue of which school improvement initiatives or strategies for improvement provide successful *levers* to the improved performance of schools over time. For example, what is the affect of providing feedback data to schools on their effectiveness? A particular issue relates to the variety of strategies that may be successful in different types of context, such as in areas of high versus low socio-economic disadvantage. The Improving School Effectiveness Project (MacBeath & Mortimore, 1994) is currently addressing these issues and findings of the project will provide important information regarding the implementation and impact of particular strategies for school improvement (Robertson *et al* 1998).

Conclusions

The results of this study provide important evidence that schools need to continuously review their internal variations in performance in any year and across years in order to monitor possible differences in the educational quality and standards for different groups of pupils, at the departmental, subject and classroom level as well as overall. The findings also support the results of previous research which showed that few schools perform both consistently across subjects and with stability over time (Thomas *et al*, 1997a). These findings are of practical as well as theoretical importance. School performance that varies greatly over time or between departments in secondary schools has implications for whole school policies and may provide important evidence about the impact of school improvement initiatives. School performance that varies greatly for different groups of pupils (such as boys and girls) has implications for equal opportunities and pupil entitlement within a school.

However, new evidence is also reported on the topic of regional differences in the range and extent of school effects which suggests that the regional, socio-economic or educational policy context of a school influences substantially the possibilities of improved performance (for example, the extent of pupil selection or private schooling in the local region). Differences in the range and extent of school effects in different UK regions suggests that socio-economic context and national and regional educational policy and practice - factors which are largely outside the control of the individual school - may play an important role in the average level of a school's effectiveness and it's opportunities for improvement. Hopefully, evidence of this kind, employed as part of a confidential framework for school and teacher self evaluation, will stimulate and inform teacher's evaluation of their own educational practices and capacity for improvement as well as the overall quality of teaching and learning in their school and the local region or LEA.

In the light of this and previous research, we would argue that effectiveness is best seen as a feature which is outcome, time and also to some extent, context specific. Therefore judgements about schools need to address at least four key questions:

◊ Effective in promoting which outcomes?

◊ Effective over what period of time?

◊ Effective for whom?

◊ Effective in what educational policy or regional context?

However, further work is required to examine in detail:

◊ Additional dimensions of school effectiveness. What outcomes of schooling are valued, in addition to those reported in the current study, and how can these be measured?

◊ What is the relationship between effectiveness at different levels within the overall education system (eg national, regional, local, school, department, classroom, individual)?

◊ The limitations of the data. How well can we control for factors outside the control of the school such as additional private tuition?

◊ What conditions appear to enhance, or alternatively form barriers, to school effectiveness?

◊ What kind of strategies, or levers, appear to improve school effectiveness?

◊ What is the long term impact of school self-evaluation processes on the quality of teaching and learning?

Acknowledgements

We are grateful to our ISEIC colleagues who have provided useful comments on a previous version of this paper. Many thanks also go to Ros Marsh for typing the manuscript and her cheerful and endless support. Finally we are indebted to all the LEA and government officers, heads and teachers who provided data for this project.

References

Coleman, J S, Campbell, E, Hobson, E, McPartland, J, Mood, A, Weinfeld, F & York, R (1966) *Equality of Educational Opportunity*, Washington: National Center for Educational Statistics.

Creemers, B P M, Reynolds, D & Swint, F E (1994) *First Results of the Quantitative Study*, paper presented at the British Education Research Association, Oxford, September 1994.

Fitz-Gibbon, C T (1991) Multilevel modelling in an indicator system, Ch 6 in S W Raudenbush & J D Willms (Eds) *Schools, Classrooms and Pupils International Studies of Schooling from a Multilevel Perspective*, San Diego: Academic Press.

Fitz-Gibbon, C (1995) *National Value Added Project. General Report: Issues to be Considered in the Design of a National Value Added System*, London: School Curriculum and Assessment Authority.

Goldstein, H (1987) *Multilevel Models in Educational and Social Research*, London: Charles Griffin & Co.

Goldstein, H (1995) *Multilevel Statistical Models (2nd Edition)*, London: Edward Arnold & New York: Halsted Press.

Goldstein, H (1997) Methods in School Effectiveness Research, *School Effectiveness & School Improvement*, 8, (4): 369-395.

Goldstein, H, Rasbash, J, Yang, M, Woodhouse, G, Pan, H, Nuttall, D & Thomas, S (1993) A Multilevel Analysis of School Examination Results, *Oxford Review of Education*, 19, (4): 425-433.

Goldstein, H & Sammons, P (1997) The Influence of Secondary and Junior Schools on Sixteen Year Examination Performance: A Cross-Classified Multilevel Analysis, *School Effectiveness & School Improvement*, 8, (2): 219-230.

Gray, J (1993) *Value-Added Approaches in School Evaluation: The Experiences of Three LEAs in England - Lessons and Challenges*, Audit Unit, HM Inspectors of Schools, The Scottish Office.

Gray, J, Jesson, D & Sime, N (1990) Estimating Differences in the Examination Performance of Secondary Schools in Six LEAs: a multilevel approach to school effectiveness, *Oxford Review of Education*, 16, (2): 137-158.

Gray, J, Jesson, D, Goldstein, H, Hedger, K & Rasbash, J (1993) *A Multi-Level Analysis of School Improvement: Changes in Schools' Performance Over Time*, paper presented at the 5th European Conference of the European Association for Research on Learning and Instruction, 3 September, Aix-en-Provence, France.

Gray, J, Jesson, D, Goldstein, H, Hedger, K & Rasbash, J (1995) A Multilevel Analysis of School Improvement: Changes in Schools' Performance Over Time, *School Effectiveness & School Improvement*, 6, (2): 97-114.

Gray, J, Goldstein, H & Jesson, D (1996) Changes and Improvements in Schools' Effectiveness: Trends over five years, *Research Papers in Education*, 11, (1): 35-51.

Gray, J, Hopkins, D & Reynolds, D (1998) *The Improving Schools Research Project*. A paper presented at The International Congress for School Effectiveness & Improvement, Manchester, January 1998.

Hill, N (1997) *Value Added and School Performance: Recent LEA Initiatives*. Slough: National Foundation for Educational Research in England and Wales, Education Management Information Exchange.

Hill, P & Goldstein H (forthcoming) Multilevel Modelling of Educational data with cross–classification and missing identification for units, *Journal of Educational and Behavioural Statistics*.

Jencks, C S, Smith, M, Ackland, H, Bane, M J, Cohen, D, Gintis, H, Heyns, B & Micholson, S (1972) *Inequality: A Reassessment of the Effect of Family and Schooling in America*, New York: Basic Books.

Joreskog, K G & Sorbom, D (1989) *LISREL 7: User's Reference Guide*, Chicago: SPSS Inc.

Luyten, J W (1994) *Stability and Malleability*, Enschede: University of Twente, Faculty of Education.

MacBeath, J & Mortimore, P (1994) *Improving School Effectiveness – A Scottish Approach*, British Educational Research Association 20th Annual Conference, September, St Anne's College, Oxford.

McPherson, A (1992) Measuring added value in schools *National Commission on Education Briefing No 1*, February 1992, London.

Madaus, G F, Kellaghan, T, Rakow, E A & King, D (1979) The sensitivity of measures of school effectiveness, *Harvard Educational Review*, 49, 207-230.

Mortimore, P, Sammons, P, Stoll, L, Lewis, D & Ecob, R (1988) *School Matters: The Junior Years*, Wells: Open Books.

Nuttall, D, Goldstein, H, Prosser, R & Rasbash, J (1989) Differential School Effectiveness, Chapter 6 in *International Journal of Educational Research*, special issue Developments in School Effectiveness Research, 13, 769-776.

Nuttall, D L, Sammons, P, Thomas, S & Mortimore, P (1992) *Differential School Effectiveness: Departmental Variations in GCSE Attainment*, ESRC award R000234130, Institute of Education, University of London.

Paterson, L & Goldstein H (1991) New statistical methods of analysing social structures: an introduction to multilevel models, *British Educational Research Journal*, 17, (4): 387-393.

Rasbash, J & Woodhouse, G (1995) *Users' Guide for MLN*, London: Institute of Education, University of London.

Raudenbush, S W (1989) The Analysis of Longitudinal, Multilevel Data, Chapter 3 in *International Journal of Educational Research*, special issue Developments in School Effectiveness Research, 13, (7): 721-740.

Robertson, P, Toal, D, MacGilchrist, B & Stoll, L (1998) *Quality Counts: Evaluating Evidence for School Improvement*, paper presented at the International Congress for School Effectiveness & Improvement, University of Manchester, January 1998.

Rowe, K J & Hill, P W (1994) *Multilevel Modelling in School Effectiveness Research: How Many Levels?* paper presented at the International Congress for School Effectiveness & Improvement, Melbourne, Australia, 3-6 January 1994.

Rutter, M, Maughan, B, Mortimore, P & Ouston, J (1979) *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*, London: Open Books.

SCAA (1994) *Value-Added Performance Indicators for Schools*, a report by the School Curriculum and Assessment Authority to the Secretary of State for Education, London.

SCAA (1995) *The Value Added National Project: General and Technical Reports*, London: SCAA.

SCAA (1996) *Baseline Assessment Draft Proposals*, London: SCAA.

SCAA (1997) *The Value Added National Project - Report to the Secretary of State*, London: SCAA.

Sammons, P, Thomas, S, Mortimore, P, Owen, C & Pennell, H (1994) *Assessing School Effectiveness: Developing Measures to put School Performance in Context*, London: Office for Standards in Education [OFSTED].

Sammons, P, Nuttall, D, Cuttance, P & Thomas, S (1995) Continuity of School Effects: A longitudinal analysis of primary and secondary school effects on GCSE performance, *School Effectiveness & School Improvement"*, 6, (4): 285-307.

Sammons, P, Mortimore, P & Thomas, S (1996) Do schools perform consistently across outcomes and areas? To appear in J Gray, D Reynolds, C Fitz-Gibbon & D Jesson (Eds) *Merging Traditions: The future of research on school effectiveness and school improvement*, London: Cassells.

Sammons, P, Thomas, S & Mortimore, P (1997) *Forging Links: Effective Schools and Effective Departments*, London: Paul Chapman, ISBN 1-85396-349-6.

Sammons, P & Smees, R (1997a) *Using Baseline Screening Information to Examine Pupil Progress Across Years 1 and 2*, report prepared for Surrey Education Services.

Sammons, P & Smees, R (1997b) *Using Baseline Screening Information to Examine Pupil Progress Across Years 4 to 6*, report prepared for Surrey Education Services.

Scheerens, J (1990) School effectiveness research and the development of process indicators of school functioning, *School Effectiveness & School Improvement*, 1, (1): 61-80.

Scheerens, J (1992) *Effective Schooling: Research, Theory and Practice*, London: Cassell.

Scheerens, J & Bosker, R (1997) *The Foundations of Educational Effectiveness*, Oxford: Elsevier Science.

Smith, D J & Tomlinson, S (1989) *The School Effect: A Study of Multi-Racial Comprehensives*, London: Policy Studies Institute.

Thomas, S (1995) Considering Primary School Effectiveness: An analysis of 1992 Key Stage 1 Results, *The Curriculum Journal*, 6, (3): 279-295..

Thomas, S, Nuttall, D & Goldstein, H (1993a) *The Guardian Survey (of A-level examination results) November 1993*, The Guardian, 30 November 1993.

Thomas, S, Nuttall, D & Goldstein, H (1993b) *Report on Analysis of 1991 Examination Results*, London: Association of Metropolitan Authorities.

Thomas, S, Pan, H & Goldstein, H (1994) *Report on Analysis of 1992 Examination Results* prepared for the AMA, London: Institute of Education, University of London.

Thomas, S & Mortimore, P (1996) Comparison of Value Added Models for Secondary School Effectiveness, *Research Papers in Education* , 11, (1): 5-33.

Thomas, S, Smees, R & McCall, J (1997) *Room for Improvement: Analysis of ISEP primary baseline measures*, paper presented at a symposium Critical Issues in School Effectiveness Research at 10th International Congress for School Effectiveness and Improvement, Memphis, Tennessee, January 1997.

Thomas, S, Sammons, P, Mortimore, P & Smees, R (1997a) Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years, *School Effectiveness and School Improvement*, 8, (2): 169-197.

Thomas, S, Sammons, P, Mortimore, P & Smees, R (1997b) Differential Secondary School Effectiveness: Examining the size, extent and consistency of school and departmental effects on GCSE outcomes for different groups of students over three years, *British Educational Research Journal*, 23, (4): 451-469.

Thomas, S & Elliot, K (1997a) *A Contextualised Analysis of the Summer 1996 Baseline Assessments for Year 3 & 5*, report prepared for The Department of Northern Ireland Raising School Standards Initiative, Belfast.

Thomas, S & Elliot, K (1997b) *A Contextualised Analysis of the Summer 1996 Baseline Assessments for Year 8 & 10*, report prepared for The Department of Northern Ireland Raising School Standards Initiative, Belfast.

Thomas, S, Smees, R, MacBeath, J, Sammons, P & Robertson, P (1998) *Creating a Value Added Framework for Scottish Schools*, report prepared for the Scottish Office Education & Industry Department, Edinburgh.

Willms, J D (1986) Social class segregation and its relationship to pupils' examination results in Scotland, *American Sociological Review*, 51, 224-241.

Willms, J D (1992) *Monitoring School Performance: A Guide for Educators*, London: Falmer.

Willms, J D & Raudenbush, S W (1989) A longitudinal hierarchical linear model for estimating school effects and their stability, *Journal of Educational Measurement*, 26, (3): 209-232.